Functional Encryption for Secured Big Data Analytics

Nilotpal Chakraborty
Center of Advanced Systems Engineering
Indian Institute of Technology Patna, India

G K Patra
Principal Scientist, CSIR- Fourth Paradigm Institute
Council of Scientific and Industrial Research
Bangalore, India

ABSTRACT
Big Data has gained a lot of interest now a days due to the enormous scope it introduces for a better understanding of the business and decision making process for an organization. There has been a wide range of researches going on around the globe on this to make the most out of it and quite a number of technical platforms are available to perform various big data analytics. Though, Scientists and organizations have been able to develop systems that can handle the big data and perform analysis on it, the security aspects remains vastly untouched. As big data is ultimately some data that are to be stored and processed in digital format, all the data security problems are applicable to it also. In this paper, possible implementation of 'Functional encryption' has been proposed towards a secured big data analytics infrastructure.

General Terms
Big Data, Security, Cryptography.

Keywords
Big data, fully homomorphic encryption, functional encryption, cryptography.

1. INTRODUCTION
Big data is a term defined for a set of unstructured, large volume of data that perhaps cannot be handled and processed by the available traditional database management systems. The reason behind this is basically due to the large set of unstructured, complex, varied forms of data that limits the processing capacity of the traditional database models. The term was first introduced by Roger Magoulas from O'Rielly media [1] to distinguish the processing capacity of complex and unstructured data by traditional databases. The notion of big data becomes clear in its definition as provided by Gartner Inc [2] in its 3V model as below—

Volume: The huge amount of data generated every single minute by various transactional processes and other data generation processes such as airline systems, online social media imposes a great deal of information to be stored, processed and retrieved. For an example, Facebook generates around 500 terabytes (TB) of data every day.

Variety: The issue of handling data and processing has now been more difficult due to the various types of data being disposed. There are text data, image data, video data, database files, emails etc. all of which makes the data processing, a tedious job. This is primarily, one of the biggest reasons why the existing database solutions are not suitable for big data analysis and we need specific solutions to handle such data.

Velocity: This basically reflects the speed at which the data is being generated and dispersed over the internet. With this era of internet and mobile computing, accessing information has become much simpler and consequently vast amount of information are also being generated from this usage. This introduces the need to manage the fast growing real time data and also to be able to process the data for analysis.

Internet represents a vast space where huge amount of data are being added every now and then and with the introduction of Cloud computing, the intensity has grown exponentially. According to a study conducted by IBM in 2012 [3], we have 2.7 Zettabytes of data in the digital world, which certainly has grown much more by now, and 90% of the data has mostly been generated for the past couple of years only.

Though there is a huge amount of data, the issue is not the storage as we can have huge amount of secondary storage capacity available with us. The issue that is being faced is the processing capacity and retrieval of data efficiently for analyzing and decision making. Apache Hadoop [4] has come up with the solution that can handle big data analytics. Hadoop is an open source platform that supports analyzing big data. Organizations and scientists around the World are putting their honest efforts in making big data analytics a success. But the point where there have been little advancements is the security aspects of big data. As it said, with great power come great responsibilities, organizations working on big data need to secure the data processing. They must adhere to the various rules and guidelines of security compliance and make sure about the data security. The following figure depicts the exponential growth of data with respect to time.

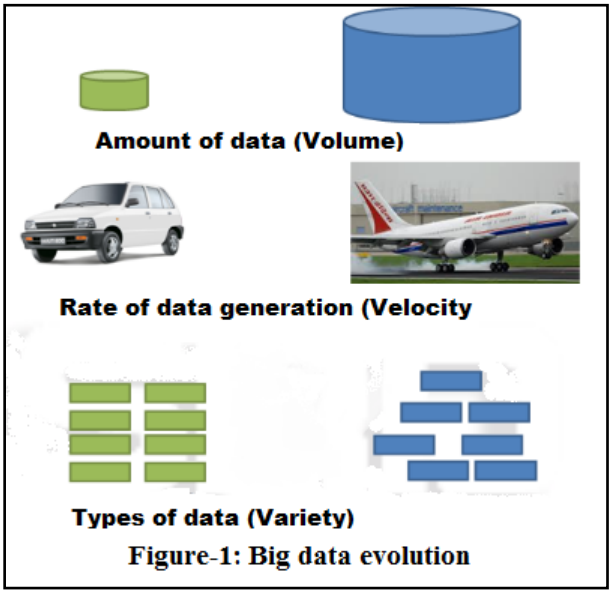

Figure-1: Big data evolution

In this paper, the primary focus is on discussing the challenges and issues related to big data analytics. The paper is structured in the following way: in section 2, the importance of big data is briefly discussed and section 3 will bring the notion of issues faced in big data analytics. In section 4 introduces Functional encryption and Section 5 will talk about its possible perspective for big data. Section 6 and 7 talks about the implementation and major contributions made in the area.

2. SIGNIFICANCE OF BIG DATA

The primary need and importance of big data can be realized by the gained efficiency in decision making process of an organization. With the enormous amount of data, organizations can analyze every minute details of their business and take decisions that meet their suitability. If big data is used properly, organizations can gain better efficiency in managing various processes such as marketing, sales etc. One of the primary beneficiaries of big data analytics would be the Government. As the Government is having huge volume of historical data, big data analytics technologies can help them in actually analyzing them and making some proper utilization of those data. Data such as birth and death records, electoral records, criminal records, financial records, railway budgets, irrigation records etc. can help in a lot of ways to the Government is making proper decisions. This can help in taking proper agenda for their upcoming activities. Big data can also help in implanting various eGovernance projects around the country, especially for India as its analysis will point out the loopholes, which need proper attention.

Apart from this, organizations that are customer-centric can make a huge impact on their business by big data. Analyzing historical data can significantly help in understanding the pattern of the customers, their interests, and various other such issues. They can serve better to their stakeholder by analyzing the data they have. Moreover, big data analysis can help in fraud detection and forensic studies and thus controlling various criminal activities. Big data can also help Engineers and Scientists in their research and improving the society. And with the improved IT infrastructure, it has become very efficient now to store and analyze the past data. Big data finds its need everywhere now as more and more analysis have become important to understand the essential elements of the World in minute details. That corresponds to data from everywhere irrespective of the demographic boundaries and individual interests.

3. CHALLENGES IN BIG DATA ANALYSIS

The primary challenge in big data analytics is the understanding of the subject. Big data, like cloud computing, means different things to different people. In order to adopt big data, organizations need to be firm clear about it and mostly there should be a concrete reasoning behind big data analytics. As we are talking about data over the Internet, it is need to be understood that not every data will have similar important value to everyone. Requirements of analyzing different sets of data will differ accordingly. As cyberspace provides a huge space to store data, there will certainly be a lot of data which are of no use. Organizations need to extract the useful data as their first step in big data analysis.

The next challenge associated with big data is the volume of data that has to be managed and processed in real time. With the technology being updated and outdated at the same time so rapidly, it becomes a tough challenge for the organizations to carry on the analysis process. In a big data world, one of the important factors is speed. Traditional analytics focus on historical data only, bid data extends this concept by including the real time data and their processing.

Apart from all the other issues, one of the issues that IT Engineers have continued to face is data security and privacy and with big data, the challenge is even much more than anticipated. Security aspects of big data have two manifolds. First, there is a need to secure user data that have been stored and analyzed. Second, organizations performing big data analytics needs to make sure that they follow the security regulations and they have authority to perform the action. Big data introduces new security challenges and concerns for the IT professionals and organizations as now they have to deal with a very huge amount of data. Organizations need to make sure that the data they have stored and analyzed are securely and the results should never be revealed to any unauthorized third party as it could lead to a heavy information leakage.

Till date, there has been little development in the security area of big data analytics and researchers were more concentrated towards developing techniques to process the vast, varied, and fast generating real time data. With the rise of Big Data and the growing ease of access to vast numbers of data records and repositories, personal data privacy is becoming ever harder to guarantee [5].

One of the primary reasons that led to this scenario is caused by online social media such as Facebook. On a single day, Facebook generated more than 500 TB of data [6], which increases occasionally during the festive seasons. With so much of data, Facebook servers are certainly one of the primary targets for the hackers and intruders. Recently there have been incidents of hacking Twitter and True caller, which eventually caused more than 2 billions of data to be in trouble. In September, 2013, Adobe was hacked where around 2.9 million of user's data stolen. As these organizations contain billions of data, access to these data can be advantageous in many ways to an adversary and therefore, strict security mechanisms are needed to mitigate these problems.

4. FUNCTIONAL ENCRYPTION

Functional encryption is a form of public key cryptography that provides a limited secret key which can encrypt a certain functionality of the encrypted data, without exploring any further details. In traditional public-key cryptosystems, the corresponding cipher text of a plain text is intended to be decrypted only by a single recipient of the encrypted data. But in certain scenario, such as cloud computing, an encrypted message may be directed to a group of people, without knowing a specific individual. In such a communication scenario, functional encryption lets a user to decrypt only a specific functionality of the cipher text, without revealing any more information about the original data. Existing encryption schemes, such as Identity based encryption and Attribute based encryption can be seen as specific cases of Functional encryption. Functional encryption was introduced in 2005 by Amit Sahai and Brent Waters [7] which supported the evaluation of some specific functionality. In 2012, several researchers around the world developed Functional Encryption schemes that support arbitrary functions. Thus it introduces a new form of cryptographic encapsulation on the cipher texts.

A functional encryption consists of the following algorithms:

(pk, msk) ← *SetUp(1):* creates a public key *pk* and a master secret key *msk.*

sk ← *KeyGen(msk,k):* uses the master secret key to generate a new secret key *sk* for value *k.*

C ← *Encrypt (pk,m):* uses the public key to encrypt a message *m.*

F(k,m) ← *Dec(sk,c):* uses secret key *sk* to calculate a function of the value *c* encrypts.

The secret key *sk* in the above construction, restricts the user from decrypting any other information apart from the

functionality for which the secret key is provided for. Thus this scheme can help in providing limited access to certain resources which must be protected and accessed from a limited set of users, though the resources reside on public servers. The rough working mechanism of functional encryption can be depicted in figure-2.

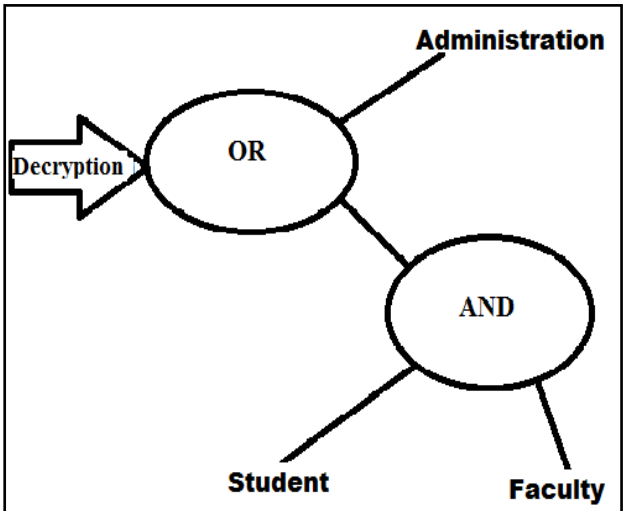

Figure-2: Working of Functional Encryption

5. FUNCTIONAL ENCRYPTION AND BIG DATA

As discussed earlier, data security in big data analytics is much more important and challenging due to the enormous amount and varied form of data coming from multiple sources. Traditionally, encryption has been considered as the best way to secure data as it converts original data to a form that is meaningless without its proper decryption. But as big data analytics will be used to perform analytics on the data that will mostly be stored in distributed systems and technology such as cloud computing will be used extensively, ordinary encryption schemes will not suffice the need as it will not allow performing any function on the cipher texts. Moreover, traditional encryption schemes are characterized with encrypt all-or-nothing, consequently reducing the power of selective decryption of cipher texts.

Functional encryption [8], on another hand, supports the group secrecy mechanism by only allowing a portion of the functionality of the cipher text to be revealed. The secret key of the scheme can only reveals specific functionality that restricts the user to learn nothing beyond their privilege. For example, suppose for a medical study, students need to conduct a survey about how many patients were infected by swine flu during the period of April, 2012 to March 2013. Now here if it is considered that the medical data of the patients are encrypted, then according to the needs of the study, only the number of the patients should be revealed, without deciphering any other information such as patient name, address or phone number. Thus Functional encryption helps in a number of ways where a set of cipher text, though can be seen by everyone, but only a specific portion of it can be decrypted and only a certain pre-specified computations may be done on them based on the secret key.

As in big data analytics, data will be essentially stored in virtual distributed servers, using cloud computing technologies; functional encryption can help in maintaining the secrecy and privacy of the data. Functional encryption is based on the concepts of fully homomorphic encryption [9] which eventually helps in performing operations on encrypted data and attributes based encryption [10] that allows encryption and decryption based on certain attributed of the cipher texts. Encrypting data in big data analytics with functional encryption can secure the data and also will allow performing operations on the data and only some specific functionality can be decrypted with the help of the secret key that is generated.

6. PROPOSED IMPLEMENTATION OF FUNCTIONAL ENCRYPTION

Functional encryption, as promised by the various researchers [11] [12], can support encryption on multiple functions, even with the user-defined ones and we have already witnessed specific cases such as Attribute based and Identity based encryption schemes. In ABE, the encryption is performed with the condition specifying the attribute that only can decrypt certain cipher text and in IBE, it is the identity. But Functional encryption is an advanced scheme, that lets a user to learn some specific functionality to learn on the cipher text, according to his functional secret key, and absolutely nothing else at all.

Functional encryption inherently needs Fully Homomorphic Encryption as ultimately it needs to perform some operations on the cipher texts based the function specified. As part of the implementation, the work started with implementing fully homomorphic encryption. The authors followed Fully homomorphic encryption construction as proposed by VanDijk et. al. [13] and validated its homomorphic characteristics; then finally moved towards the Functional encryption construction as devised in [14] to implement it.

The implementation was done in two parts: The first part is the Fully Homomorphic encryption (FHE) part and the second part is the Functional encryption part. In the FHE part, the developed system asks plain texts to be entered by the user and performs encryption of those plain texts to obtain cipher texts. As it follows FHE, the cipher texts themselves can be operated on. But to allow specific operations and computations on them, the Functional encryption part assigns specific functionality to the user secret keys with which only the specified computations can be carried on the cipher texts. Any secret key that was not created for any particular function can never decrypt or carry out that function on the cipher texts. It can only decrypt the functionality on the cipher texts for which it was originally created during the Key Generation stage.

All the developments and implementations of the encryption schemes are done in C programming language with the help of some open source libraries as and when required. C libraries such as GNU Multi Precision Library, Openssl library, PBC library have been used as per the requirements. All the programs are tested on Linux based systems with Intel processors.

7. MAJOR CONTRIBUTIONS

Scientists at CSIR- Fourth Paradigm Institute, Bangalore have successfully identified the importance of big data analytics and are carrying out researches on implementing big data analytics on various inter-disciplinary areas such as climate modeling, crop modeling, environmental modeling, rain forecasting, cyclone prediction etc. To perform all these scientific and engineering modeling, High Performance Computing (HPC) infrastructure is used that handles large amount of scientific data. These data used for various models

to take crucial decisions and scientific forecasting. Security of such data poses significant importance to the scientific community and as well as to the industry experts.

The HPC infrastructure available in CSIR Fourth Paradigm Institute, Bangalore is used by various Scientists for various modeling purposes and users from multiple locations are accessing the system. It is very important to protect the privacy and security of the data for its proper and appropriate implementation. Through this work on Functional encryption, the authors identified a number prominent area where such a cryptographic technique can mitigate the security problems, such as in big data analytics. Study and analysis have been done on various security issues and challenges related to big data analytics and identified such prominent issues that can be addressed with the help of cryptographic schemes. All the discussed schemes are implemented in C programming and currently, effort is being made to modify the codes to enable parallel processing.

In this work, various fully homomorphic encryption schemes have been implemented, starting from the Gentry's blueprint [15]. The inefficiencies of the Gentry's scheme were soon realized and work has been carried out on some of the other schemes available on the literature. Complete study and analysis of fully homomorphic encryption schemes have been done and it can be inferred that a practical FHE implementation is yet to be developed. The authors then focused on the applications that can be addressed with the help of somewhat homomorphic encryption schemes that would require a limited number of functionalities to be performed on the cipher texts such as image processing, video processing. Finally, some work has been carried out on functional encryption with the target to implement it on the HPC infrastructure. The implementation, though still is in the initial stage, can significantly increase the security of various security issues related to big data analysis.

8. CONCLUSION

Big data is a term for those sets of complex and huge volume data that can overwhelm the processing capability of traditional database systems or transaction processing systems. Though a number of systems and technologies have evolved till data to manage and process the big data, security issues and challenges remain to be rarely unaddressed. In this project work, the primary challenge that organizations in big data analytics face i.e., the challenge of data security has been addressed. Hereby proposed the implementation of Functional Encryption for secured big data analytics, as a powerful asymmetric key encryption technique that can eventually led the user to encrypt and access only a limited functionality of the plain text, without revealing any further details. Functional encryption, though still in its infancy can significantly lead to a highly secured digital environment for data analytics. If properly implemented, functional encryption can significantly reduce all the security challenges that Big Data analytics may face. Although, construction of an efficient functional encryption scheme remains to be a major research work in the field of cryptography.

9. ACKNOWLEDGEMENTS

Nilotpal is thankful to the SPARK program of CSIR- Fourth Paradigm Institute, Bangalore for giving him the opportunity to carry out his major research project during M.Tech in the organization. The work is partially supported by the project ARiEES, funded by CSIR, India, under the 12th Five Year Plan.

10. REFERENCES

[1] Discussion on Big data by Roger Magoulas, http://strata.oreilly.com/2010/01/roger-magoulas-on-big-data.html

[2] Gartner, Big Data Definition, http://www.gartner.com/it-glossary/big-data/, accessed in September, 2013

[3] A comprehensive list of Big Data Statistics, http://www.wikibon.rg/blog/big-data-statistics/

[4] Apache Hadoop homepage, http://www.hadoop.apache.org, accessed in September, 2013.

[5] The White Book of Big Data: The definitive guide to the revolution of business analytics, Fujitsu Services Ltd., July 2013

[6] http://www.slashgear.com , accessed in August 2013

[7] Amit Sahai, Brent Waters, Fuzzy Identity-Based Encryption, Proceedings of Eurocrypt 2005.

[8] Dan Boneh, Amit Sahai, Brent Waters, Functional Encryption: Definitions and Challenges, Proceedings of Cryptography Conference, 2011.

[9] Craig Gentry. A fully homomorphic encryption scheme. PhD thesis, Stanford University, 2009. http://crypto.stanford.edu/craig

[10] Vipul Goyal, Omkant Pandey,Amit Sahai, Brent Waters, Attribute Based-Encryption for Fine-Grained Access Control of Encrypted Data, ACM CCS, 2006.

[11] S. Agarwal, S Gorbunov, V. Vaikuntanathan, H. Wee; Functional Encryption: New Perspective and Lower Bounds, Cryptology ePrint Archive, Report 2012/468, Available online: www.eprint.iacr.org/2012/468

[12] S. Agarwal, S. Agarwal, S. Badrinarayanan, A. Kumarasubramanian, M. Prabhakaran, A. Sahai; Function Private Functional Encryption and Property Preserving Encryption: New Definitions and Positive Results,Cryptology ePrint Archive, Report 2013/744, Available online: www.eprint.iacr.org/2013/744

[13] Marten van Dijk, Craig Gentry, Shai Halevi, and Vinod Vaikuntanathan. Fully homomorphic encryption over the integers. In Advances in Cryptology - EUROCRYPT'10, volume 6110 of Lecture Notes in Computer Science, pages 24-43. Springer, 2010. Full version available on-line from http://eprint.iacr.org/2009/616

[14] S Goldwasser, Y Kalai, R Popa, V Vaikuntanathan, N Zeldovich; Reusable Garbled Circuits and Succinct Functional Encryption, Cryptology ePrint Archive, Report 2012/733, Available online: https://eprint.iacr.org/2012/733

[15] C. Gentry. Fully homomorphic encryption using ideal lattices. In *STOC '09*, pages 169-178, ACM, 2009.